Microsoft Innovation & Policy Center
901 K Street, NW 11th Floor
Washington, DC 20001

Tel 202-263-5946
Fax 202-783-0583
Paula.Boyd@microsoft.com

Microsoft

9 August 2021

<u>*Via ECFS*</u>
Marlene H. Dortch, Esq.
Secretary
Federal Communications Commission
445 12th Street, SW
Washington, DC 20554

> Re: *Establishing the Digital Opportunity Data Collection,* WC Docket No. 19-195;
> *Modernizing the FCC Form 477 Data Program*, WC Docket No. 11-10; *Addressing the*
> *Homework Gap through the E-Rate Program*, WC Docket No. 21-31; *Establishing*
> *Emergency Connectivity Fund to Close the Homework Gap*, WC Docket No. 21-93.

Dear Ms. Dortch:

On Friday, 6 August, the undersigned, Gunnar Halley, John Kahan, Allen Kim, and Vickie Robinson – all
from Microsoft Corporation – met via video conference call with Acting Chairwoman Jessica
Rosenworcel and Trent Harkrader, Acting Special Advisor to the Chairwoman, to discuss the
Commission's efforts to achieve higher quality mapping of broadband availability and usage across the
United States. The Microsoft participants used the attached PowerPoint deck in their presentation, and
on Thursday, 5 August, in preparation for the call, the undersigned shared a copy of the deck via email
with Trent Harkrader.

Microsoft provided data analysis estimating that in October 2020 approximately 120.4 million people in
the United States – more than a third of the U.S. population – were not using the internet at broadband
speeds (greater than or equal to 25 Mbps). Microsoft also estimates that there were roughly 19 million
children in households not using the internet at broadband speeds during the height of the pandemic
lockdown. We explained that broadband usage and availability while not the same are highly
correlated, and described how to use machine learning models to predict broadband availability and
improve mapping quality. Finally, Microsoft emphasized that using differential privacy approaches, one
can make available big data sets for analysis in a non-trusted environment without compromising
individual privacy. For example, the Commission could use SmartNoise, the differential privacy open
platform developed by Microsoft and Harvard Institute for Quantitative Social Science, to open up new
mapping data as it becomes available.

Pursuant to the Commission's rules, I have filed a copy of this notice electronically in the above-referenced dockets. Please contact me if you require any additional information.

Respectfully submitted,

/s/ Paula Boyd

Paula Boyd
Senior Director, U.S. Government Affairs

cc: Trent Harkrader

Broadband mapping meeting with Chairwoman Rosenworcel

John Kahan, VP, Chief Data Analytics Officer
**August 6, 2021**

Objective: ensure we are doing whatever we can to assist you on your journey to get the highest quality broadband maps
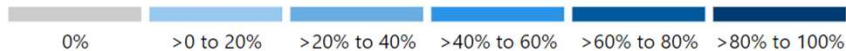
- Agenda
  - Review Microsoft's latest broadband mapping assets that can be used to assist now and in the future
    - Review FCC broadband availability and broadband usage based on Microsoft data
      - State, county, zip code, and congressional district
    - Review the analysis of children in households not using the internet at broadband speeds
    - Potential additional areas we can help with technology and AI

  - Questions:
    - What is the plan on mapping and funding allocations between now and when the new availability maps are completed?
    - How long do you think it will take to deliver a usable version of the new broadband availability maps?
    - Are you considering during the transition period and the final version of your broadband maps to use multiple data sources to improve quality and the funding allocation process?

  - Discussion

FCC indicates broadband is not available to ~14.5M (~4.4%) people
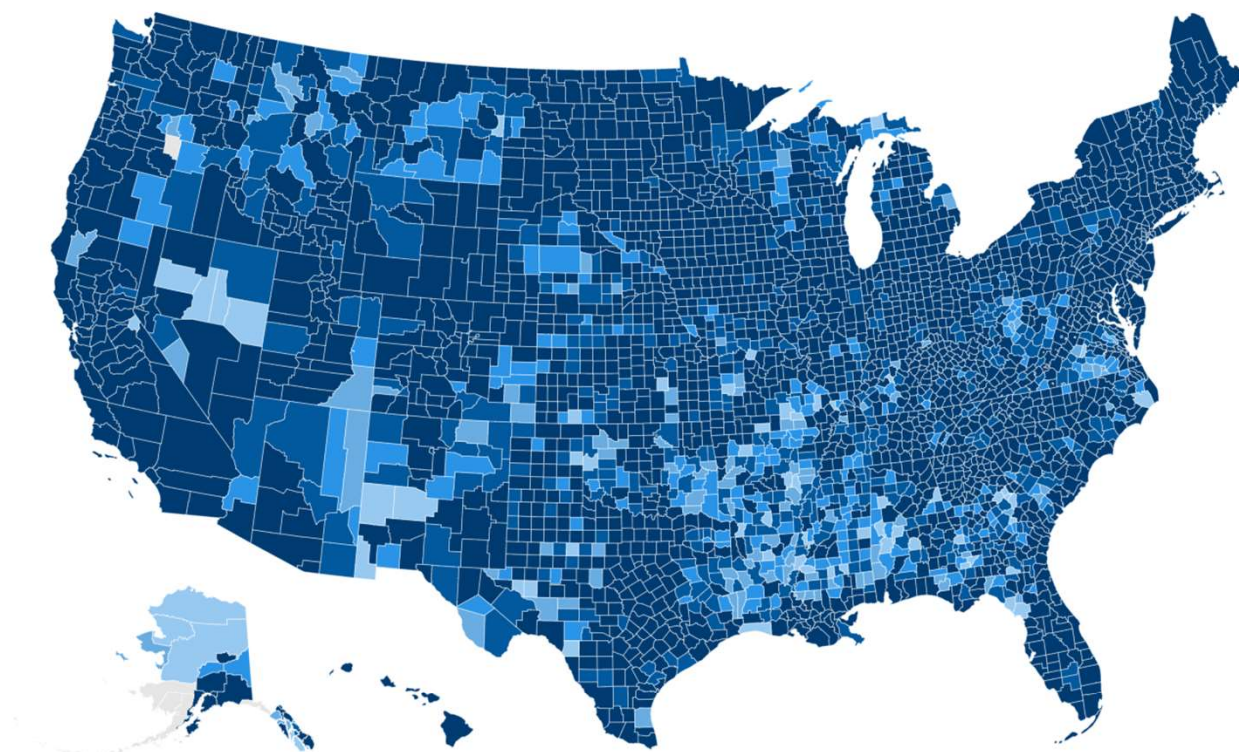
Microsoft

Select a View

FCC broadband availability

FCC and Microsoft

Congressional districts

Broadband subscriptions

Select a State

All

FCC broadband availability*

0%   >0 to 20%   >20 to 40%   >40 to 60%   >60 to 80%   >80 to 100%

* FCC Broadband has or "could" provide greater than or equal to 25 Mbps / 3 Mbps

Sources: FCC Fourteenth Broadband report based on form 477 data from December 2019 and Microsoft data from October 2020
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.

Link to Broadband dashboard

3

Comparison of broadband calculation methods

Current method using Form 477

- Internet service providers submit data by census block

- Available: For purposes of this form, fixed broadband connections are available in a census block if the **provider does, or could, within a service interval that is typical for that type of connection**—that is, without an extraordinary commitment of resources…"

- Not clear on how FCC verifies data and provides a method for the public to challenge accuracy of the maps

Census blocks (~11M national) ⟹ Census blocks submitted via csv in the format below (e.g. 511790104061009 is a census block id)

Block Code	DBA	Tech Code	Consumer Flag	MaxAdDn (Mbps)	MaxAdUp (Mbps)	Business Flag
511790104061009	AAAInternet	42	1	12	1.5	1
511790104061010	AAAInternet	42	1	12	1.5	0
511790104061011	AAAInternet	42	0	0	0	1

aaa_internet_fbd.csv - Notepad
File Edit Format View Help
511790104061009,AAAInternet,42,1,12,1.5,1
511790104061010,AAAInternet,42,1,12,1.5,0
511790104061011,AAAInternet,42,0,0,0,1

Form 477 Instructions link

New Broadband Data Act

- Broadband DATA Act requires the Commission, within 180 days of its enactment, to issue rules to:
  1. require the semiannual collection and dissemination of granular data relating to the availability and quality of service of fixed and mobile broadband Internet access service for use in conjunction with **creating broadband coverage maps**
  2. establish processes for the Commission to verify and protect the data collected
  3. establish a process for collecting verified data for use in the coverage maps from State, local, and Tribal governmental entities, from other federal agencies, and, if the Commission deems it in the public interest, from third parties
  4. establish the **Fabric to serve as a foundation on which fixed broadband availability is overlaid**
  5. establish a user-friendly challenge process through which the public and State, local, and Tribal governmental entities **can challenge the accuracy of the coverage maps**, provider availability data, or information in the Fabric;21
  6. develop a process through which entities or individuals may submit specific information about the deployment and availability of broadband Internet access service in the United States on an ongoing basis.2
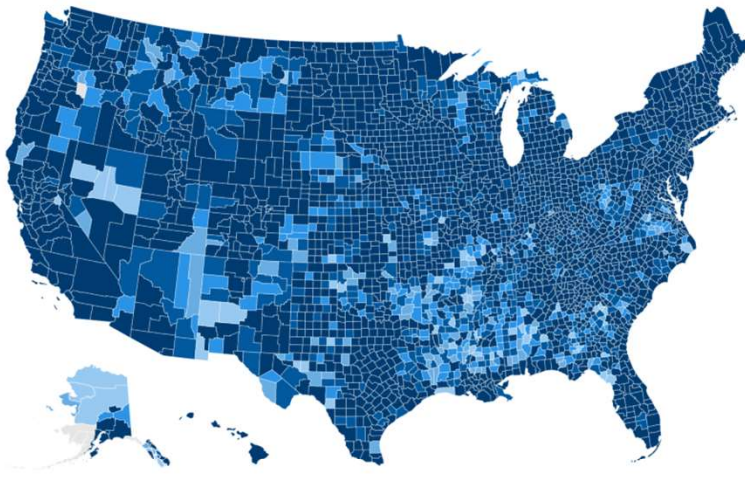
ISPs submit coverage maps ⟹ Coverage maps are overlaid on the "Fabric"
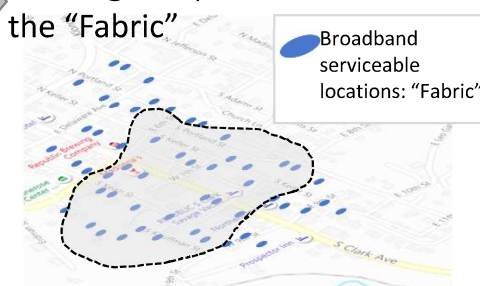
Broadband serviceable locations: "Fabric"

Maps showing FCC fixed broadband availability and broadband usage based on Microsoft data updated as of October 2020

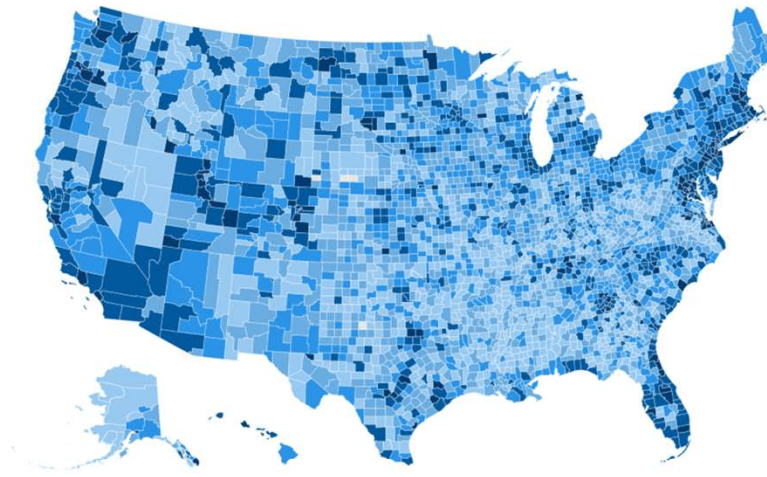FCC indicates broadband is not available to ~14.5M (~4.4%) people

Microsoft data indicates ~120.4M (36.7%) people do not use the internet at broadband speeds

Select a View

FCC broadband availability

FCC and Microsoft

Congressional districts

Broadband subscriptions

Select a State

All

FCC broadband availability*

| 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100% |

* FCC Broadband has or "could" provide greater than or equal to 25 Mbps / 3 Mbps

Broadband usage**

| 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100% |

** Broadband speeds greater than or equal to 25 Mbps

Link to Broadband dashboard

Sources: FCC Fourteenth Broadband report based on form 477 data from December 2019 and Microsoft data from October 2020
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.

5

# Using machine learning to predict broadband availability

- We use a random forest model and measure variable importance.
- Broadband usage is the #1 variable with the highest predictive power followed by percent of farmer and educational attainment.
- At the state level this model can predict with an r^2 of 91%.

Predicted Broadband vs FCC broaband (State Level)

$R^2 = 0.91$

Variable Importance
IncNodePurity
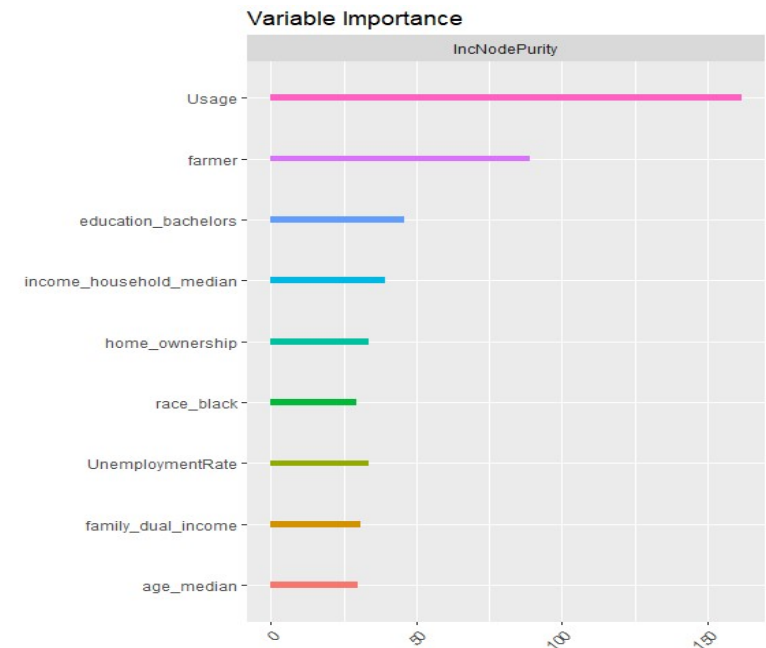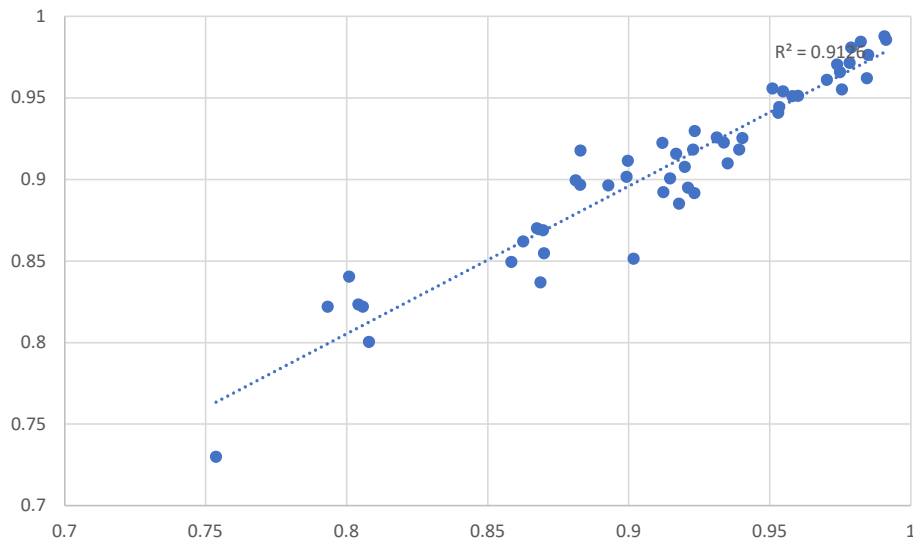
We use this model to detect potential outliers with the highest divergence (positive and negative) to the reported broadband availability.

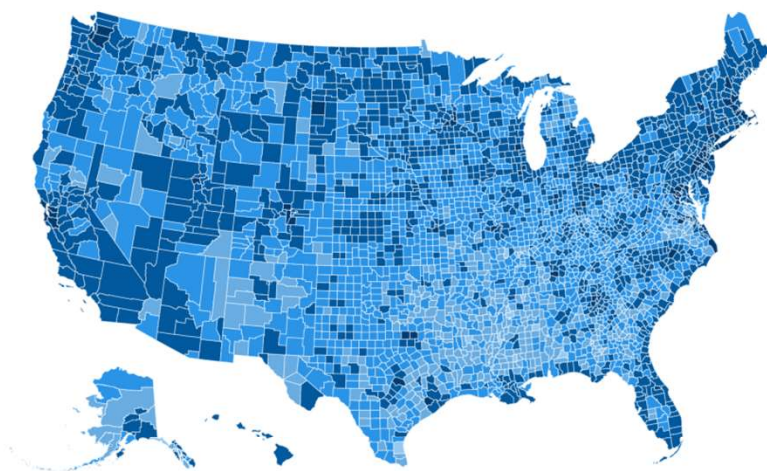- ML model predicts 78% of zip codes within 10pts of reported broadband availability

FCC reported availability data compared to ML model
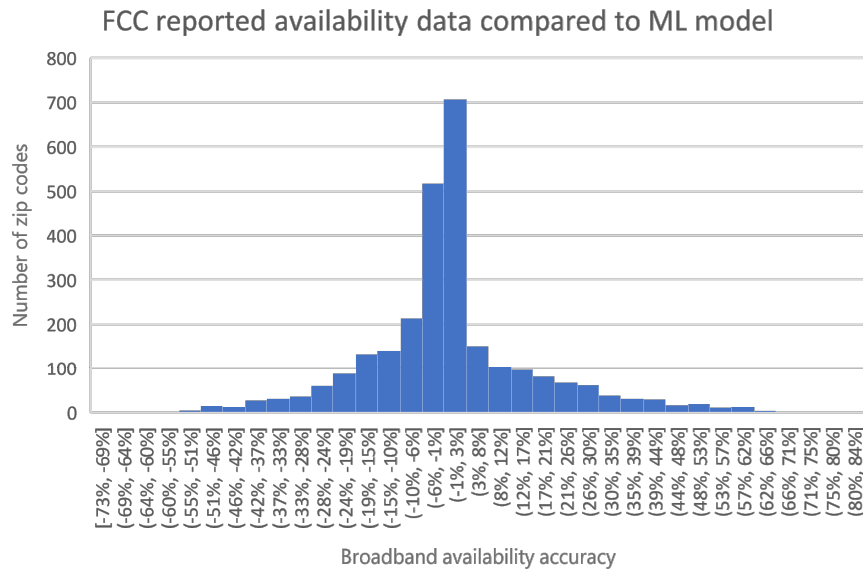
Top 20 Zip Codes

State	Zip code	FCC broadband availability 2019 report	Usage Feb 2019
PA	17949	91.8%	0.0%
VA	22742	100.0%	0.5%
WV	26386	100.0%	9.5%
FL	33890	94.0%	4.7%
OH	44076	92.1%	5.9%
OH	45856	98.3%	4.5%
IA	50514	98.0%	4.5%
MN	56282	100.0%	3.7%
KS	66079	100.0%	0.6%
AR	71956	97.3%	6.1%
AR	71968	99.1%	7.7%
OK	74332	99.9%	0.7%
TX	78118	100.0%	3.4%
TX	78151	99.6%	0.5%
TX	78941	99.5%	2.8%
CA	93602	93.7%	8.5%
CA	95638	100.0%	2.2%
OR	97456	94.7%	7.6%
WA	98855	97.9%	7.4%
WA	99122	100.0%	4.0%

Microsoft

Maps showing households with a broadband subscription from the American Community Survey and broadband usage based on Microsoft data updated as of October 2020

American Community Survey estimates ~68.9% of households have a broadband subscription (such as cable, fiber optic or DSL)
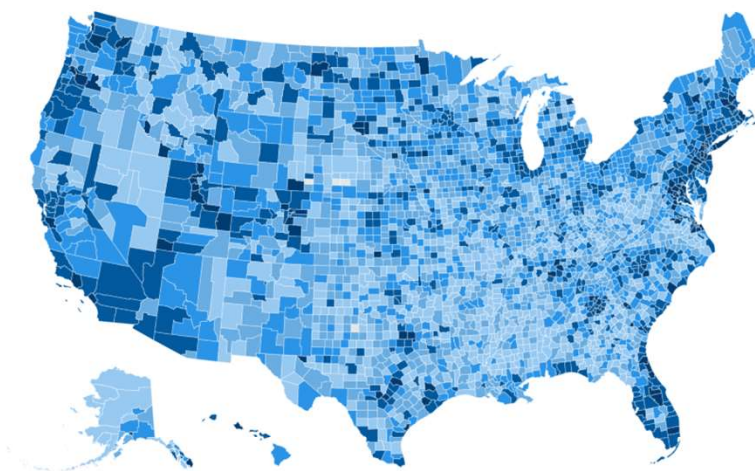
Microsoft data indicates ~63.3% of people use the internet at broadband speeds

Select a View

FCC broadband availability

FCC and Microsoft

Congressional districts

Broadband subscriptions

Select a State

All

Households with broadband subscription*

| 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100% |

* Households with broadband subscriptions such as cable, fiber optic or DSL - excludes cellular data plans or satellite internet service

Broadband usage**

| 0% | >0 to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100% |

** Broadband speeds greater than or equal to 25 Mbps

Link to Broadband dashboard

Sources: 2019 American Community Survey 5 year estimate TableID: S2801 and Microsoft data from October 2020.
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.
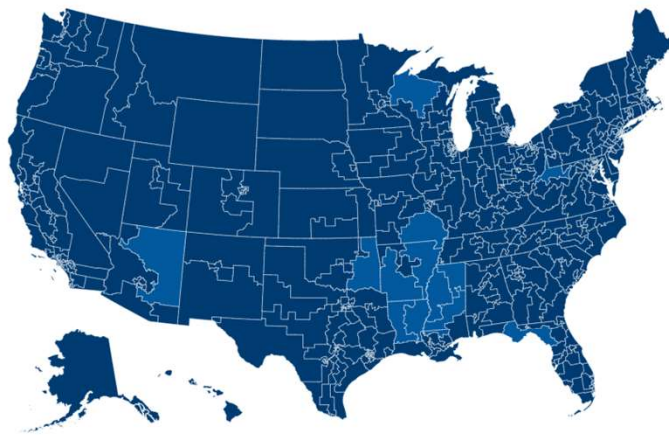
8

Broadband data made available by congressional district to enable Members of Congress to track the progress we are making to close the broadband gap

Microsoft Maps showing FCC fixed broadband availability and broadband usage based on Microsoft data updated as of October 2020

FCC indicates broadband is available to ~95.6% of people

Microsoft data indicates ~63.3% of people use the internet at broadband speeds

Select a View
- FCC broadband availability
- FCC and Microsoft
- Congressional districts
- Broadband subscriptions

Select a State

All

Select a Congressional District

All

[Link to Broadband dashboard](#)

FCC broadband availability*   0%   >0 to 20%   >20% to 40%   >40% to 60%   >60% to 80%   >80% to 100%

* FCC Broadband has or "could" provide greater than or equal to 25 Mbps / 3 Mbps
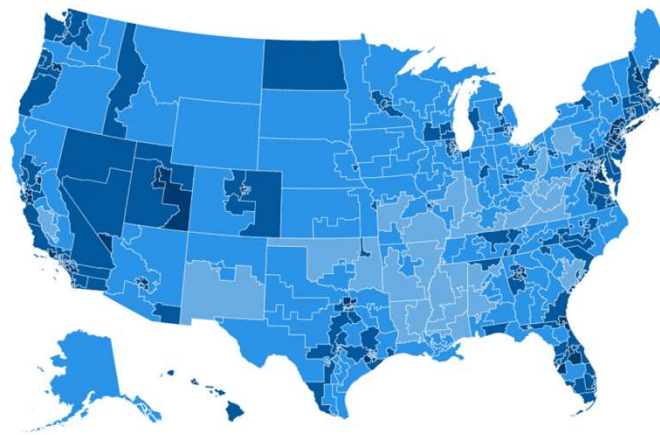
Broadband usage**   0%   >0 to 20%   >20% to 40%   >40% to 60%   >60% to 80%   >80% to 100%

** Broadband speeds greater than or equal to 25 Mbps
Zip code level data used may overlap multiple congressional districts
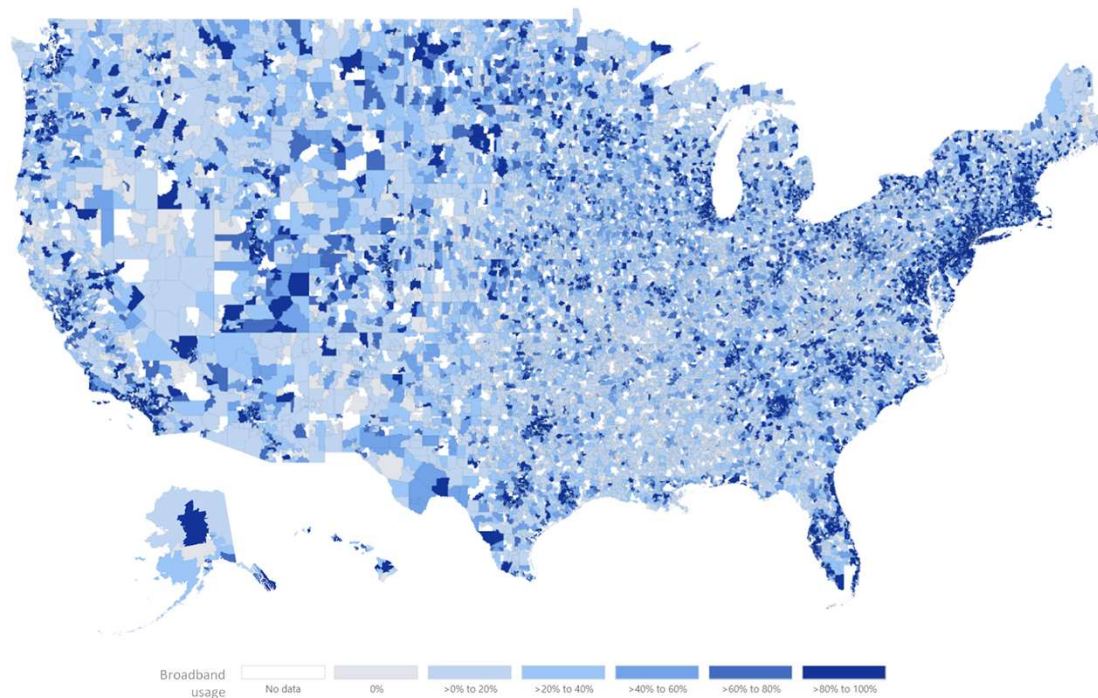
Sources: FCC Fourteenth Broadband report based on form 477 data from December 2019 and Microsoft data from October 2020
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.

Microsoft has provided publicly available broadband usage data at a zip code level on GitHub.

Broadband usage per zip code – differential privacy

Broadband usage | No data | 0% | >0% to 20% | >20% to 40% | >40% to 60% | >60% to 80% | >80% to 100%

Broadband speeds greater than or equal to 25 Mbps

Zip code list example

| ST	COUNTY NAME	COUNTY ID	POSTAL CODE	BROADBAND USAGE	ERROR RANGE (95%)(+/-)
OH	Athens	39009	45711	0.0100	0.086
KY	Bell	21013	40845	0.0100	0.136
TN	Benton	47005	38341	0.0100	0.043
MI	Calhoun	26025	49029	0.0100	0.072
NH	Carroll	33003	3883	0.0100	0.608
NY	Cattaraugus	36009	14065	0.0100	0.061
SC	Charleston	45019	29426	0.0100	0.072
MS	Coahoma	28027	38720	0.0100	0.110
AR	Conway	5029	72027	0.0100	0.072
OH	Darke	39037	45362	0.0100	0.072
NM	Dona Ana	35013	87940	0.0100	0.170
TX	Eastland	48133	76470	0.0100	0.031
MT	Flathead	30029	59919	0.0100	0.136
WV	Gilmer	54021	25267	0.0100	0.136
VA	Greensville	51081	23867	0.0100	0.043
OK	Jackson	40065	73556	0.0100	0.406
PA	Jefferson	42065	15715	0.0100	0.170
MN	Kandiyohi	27067	56216	0.0100	0.136
SC	Kershaw	45055	29175	0.0100	0.228
AK	Kodiak Island Borough	2150	99644	0.0100	0.406

Link to broadband data on GitHub

National Summary

- FCC shows an improvement of 1.2 percentage points from the 2020 report to 2021 report (95.6%)
- Broadband usage based on Microsoft data indicates ~11 percentage point improvement from Nov. 2019 to Oct. 2020
- 2019 Census data shows ~68.9% of households have a broadband subscription such as cable, fiber optic or DSL
- High correlation (~.87) between broadband subscription and broadband usage
- There are ~~19M children in households not using the internet at broadband speeds

Sources		Previous Report	Latest Report	Change
Broadband availability from the FCC	Source Reports | Federal Communications Commission (fcc.gov)	2020 Report Form 477 data from December 2018	**14th Broadband Deployment** Report (2021 Report) Form 477 data from December 2019	
	% Availability (25 Mbps / 3Mbps)	94.4%	**95.6%**	Increase of ~1.2 % pt.
	Population WITHOUT access	~18.3M	~14.5M	Decrease of ~3.8M
	Population in rural areas WITHOUT access	~14.4M	~11.3M	Decrease pf ~3.1M
Broadband usage based on Microsoft data	Source October 2020	November 2019	**October 2020**	
	% of population using the internet at broadband speed (25Mbps)	~52%	**~63%** (~207.8M people)	Increase of ~11 % pt.
	Population NOT using the internet at broadband speed	~157.3M Link to EXTERNAL Power BI	~120.4M Link to Broadband - Power BI	Decrease of ~37M
	Children in households NOT using the internet at broadband speeds	-	~19M	-
Broadband subscription from the Census	Source Census TableID: S2801	2018 ACS 5 Year Estimate	**2019 ACS 5 Year Estimate**	
	Broadband subscription of any type	~80.4% (~96.3M households)	~82.7% (~99.9M households or ~271.4M people)	Increase ~2.3 % pt.
	Broadband subscription such as cable, fiber optic or DSL (excludes cellular data plan and satellite)	~67.9% (~81.3M households)	**~68.9%** (~83.2M households or ~226.1M people)	Increase ~1.0 % pt.

# 2020 Students (K-12) not using the internet at broadband speeds at home

- As part of COVID-19 lockdowns, majority of states have a significant portion of their students with online classes.
- In order to be able to participate in online classes, students needs to have access to internet and ideally broadband connection at home.
- We want to estimate how many school children would have access to broadband at home.
- Even though many states don't have state orders in effect for partial or full closure in effect , over 65% of students are part of distance learning using online resources [1]

Where schools are reopening in the US: March 2021

Percentage of households with children reporting the use of online distance learning: Oct 2020
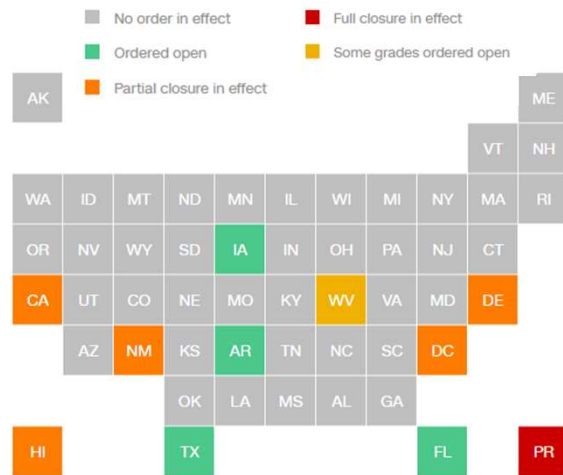
Source: Map: Where schools are reopening in the US (cnn.com)
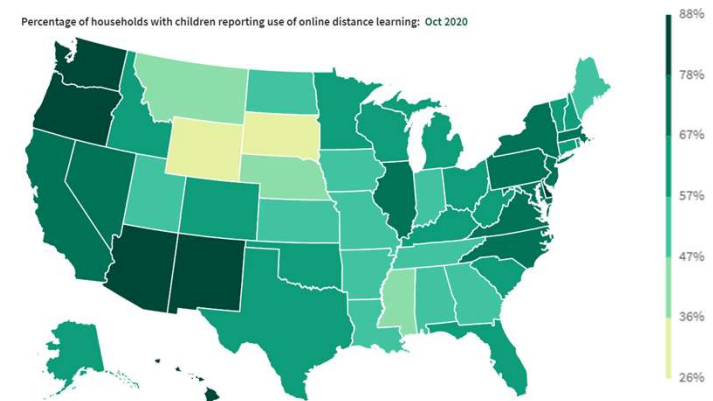
[1] 65% of households with children report the use of online learning during pandemic - USAFacts

12

We estimate in the US, there are ~19M children in households not using the internet at broadband speeds (25 Mbps / 3 Mbps)
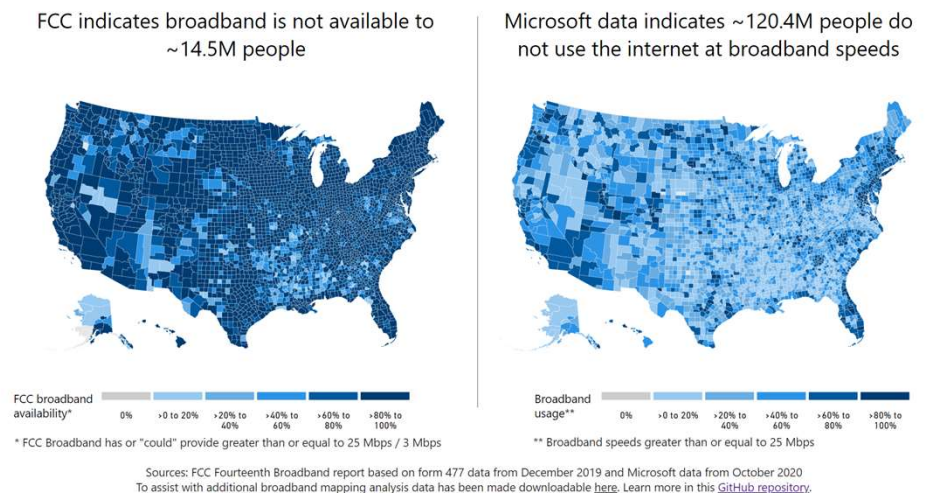

Children with broadband and without broadband

■ With Broaband ☐ No Broadband
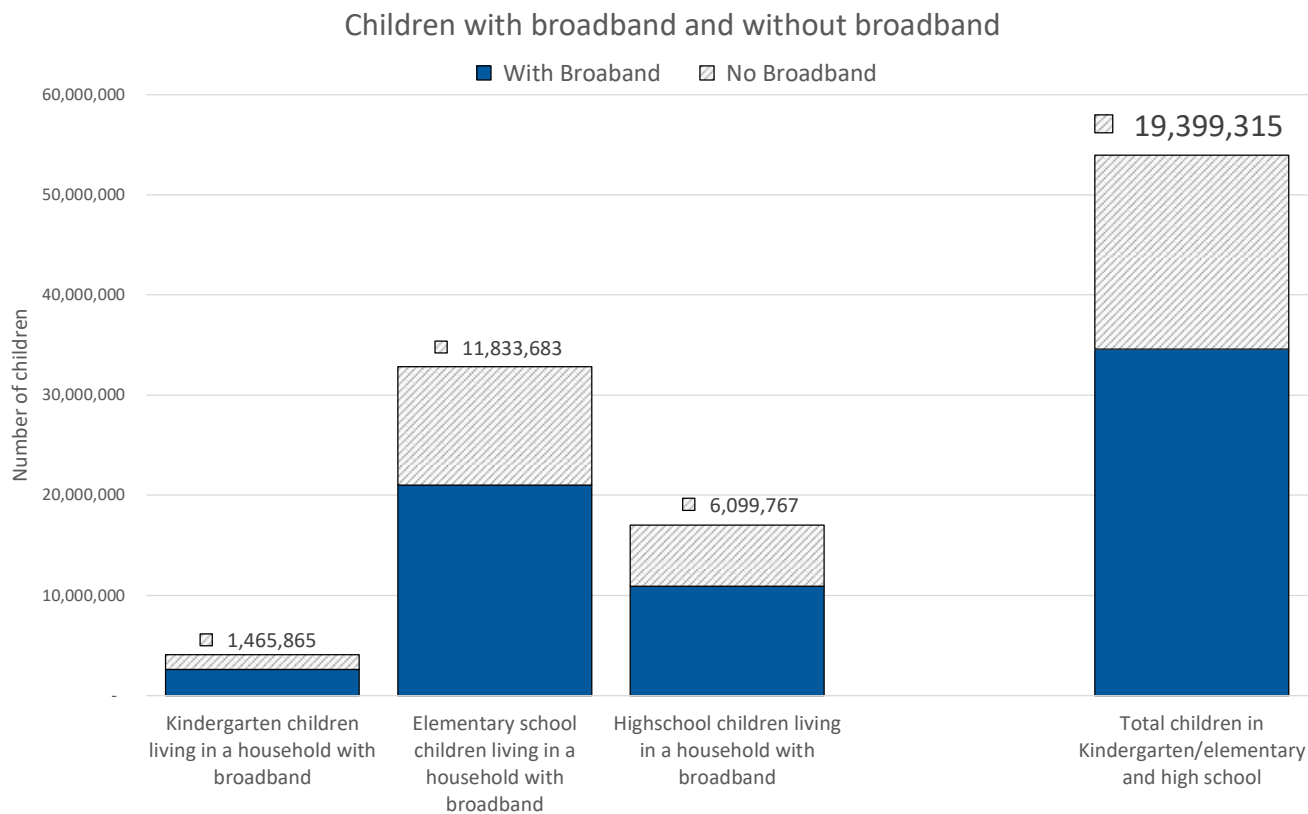
	With Broadband	No Broadband
Total children in Kindergarten in the US (census)	4,088,628	
Total children in Elementary school in the US (census data)	32,849,184	
Total children in High school in the US (census data)	17,027,129	
Total	**53,964,941**	
Kindergarten children living in a household with broadband	2,622,763	1,465,865
Elementary school children living in a household with broadband	21,015,501	11,833,683
Highschool chidlren living in a household with broadband	10,927,362	6,099,767
Total children in Kindergarten/elementary and high school	34,565,626	**19,399,315**

[1] It's time for a new approach for mapping broadband data to better serve Americans - Microsoft On the Issues
[2] GitHub - microsoft/USBroadbandUsagePercentages:

Methodology to estimate children without broadband

- We use Microsoft services[1] data to estimate how many households use internet at broadband speed. These estimates are done at the zip code level using differential privacy to provide mathematical guarantees for the user's privacy[2]

- We combine this data with census data at the zip code level. This process assume that data within the zip code is independent of the broadband connectivity within that zip code. (e.g. if a zip code has 100 people and 20 are students and the zip code has 50% using the internet at broadband speeds then we estimate 10 students use the internet at broadband speeds.)

FCC indicates broadband is not available to ~14.5M people

Microsoft data indicates ~120.4M people do not use the internet at broadband speeds

FCC broadband availability* 0% >0 to 20% >20% to 40% >40% to 60% >60% to 80% >80% to 100%

* FCC Broadband has or "could" provide greater than or equal to 25 Mbps / 3 Mbps

Broadband usage** 0% >0 to 20% >20% to 40% >40% to 60% >60% to 80% >80% to 100%

** Broadband speeds greater than or equal to 25 Mbps

Sources: FCC Fourteenth Broadband report based on form 477 data from December 2019 and Microsoft data from October 2020
To assist with additional broadband mapping analysis data has been made downloadable here. Learn more in this GitHub repository.

Census data ——— Report ——— Microsoft broadband estimates

Sources:
[1] It's time for a new approach for mapping broadband data to better serve Americans - Microsoft On the Issues
[2] GitHub - microsoft/USBroadbandUsagePercentages:
Latest available census data

Potential areas we can help

- Microsoft open data sets available on GitHub to help check data quality of new mapping data
    - State, county, and congressional district Microsoft usage data
    - Zip code level Microsoft usage data

- Machine learning algorithms can detect differences in availability reporting
    - Algorithm to be publicly available via GitHub

- SmartNoise platform can open up more granular data in a privacy compliant way to improve quality and transparency of future mapping deliverables
    - In partnership with Harvard IQSS and OpenDP

- Skills transfer on AI and Power BI visualization of new maps when they become available

- Other?

Appendix

Why do we need Differential Privacy?

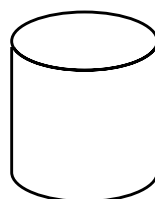"Anonymized data isn't", Cynthia Dwork (Microsoft Research)

Anonymized Movie Rating Dataset

NETFLIX

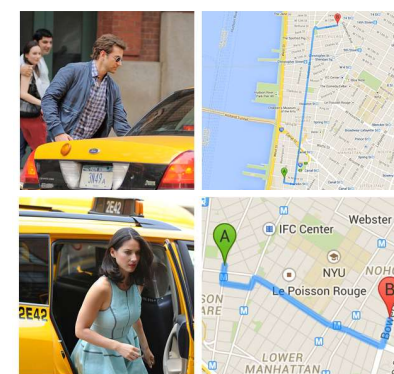| Title	Date of rating	Rating
Alien	05/02/04	★★★★
Beauty and the Beast	04/30/05	★★
Fight Club	09/23/05	★★★★
Ghostbusters	11/03/03	★★★
The Godfather	01/19/05	★★★★★
Thelma & Louise	07/30/03	★★★★
Good Will Hunting	05/08/04	★★★★
...	...	...

Internet Movie DB

IMDb

Re-identification

New York City Taxi Dataset

Paparazzi Photos with Metadata

Re-Identification

Differential Privacy & SmartNoise

Differential Privacy
Allows for statistics and/or machine learning whereby the contribution of individual data remains hidden.

SmartNoise
Open-Source platform for the implementation of differential privacy (Microsoft and Harvard University)

https://github.com/opendifferentialprivacy

End